2

Molecular Cloning

A LABORATORY MANUAL

SECOND EDITION

J. Sambrook
UNIVERSITY OF TEXAS SOUTHWESTERN MEDICAL CENTER

E.F. Fritsch
GENETICS INSTITUTE

T. Maniatis
HARVARD UNIVERSITY

Cold Spring Harbor Laboratory Press
1989

Molecular Cloning

A LABORATORY MANUAL

SECOND EDITION

Printed in the United States of America

9 8 7 6 5 4 3 2

Book and cover design by Emily Harste

Cover: The electron micrograph of bacteriophage λ particles stained with uranyl acetate was digitized and assigned false color by computer. (Thomas R. Broker, Louise T. Chow, and James I. Garrels)

Cataloging in Publications data

Sambrook, Joseph
Molecular cloning : a laboratory manual / E.F. Fritsch, T. Maniatis—2nd ed.
p. cm.
Bibliography: p.
Includes index.
ISBN 0-87969-309-6
1. Molecular cloning—Laboratory manuals. 2. Eukaryotic cells--Laboratory manuals. I. Fritsch, Edward F. II. Maniatis, Thomas III. Title.
QH442.2.M26 1987
574.87'3224—dc19 87-35464

All Cold Spring Harbor Laboratory Press publications may be ordered directly from Cold Spring Harbor Laboratory, Box 100, Cold Spring Harbor, New York 11724. Phone: 1-800-843-4388. In New York (516)367-8423.

CREATING MANY MUTATIONS IN A DEFINED SEGMENT OF DNA

At present, it is impossible to predict with accuracy the effect of substituting one amino acid for another in a protein. Current attempts to "improve" the properties of a protein therefore depend on analyzing large numbers of variants that are created by site-directed mutagenesis in promising regions (e.g., in and around the active site of an enzyme). Clearly, the number of potential variations that can be created, even in a circumscribed region of a protein, is extremely large. For example, 114 different mutants would be required simply to insert every possible amino acid at each of six locations in a protein. This number grows to 6^{19} if such substitutions are made in a combinatorial fashion. When planning this type of mutagenesis, careful choices must therefore be made to keep the numbers of mutants within manageable limits. For example, the numbers of potential mutants can be markedly reduced by avoiding replacements that are (1) highly conservative (i.e., the substitution of one amino acid with another whose chemical properties are very similar), (2) highly radical (i.e., replacing an amino acid with another whose chemical properties are completely different), or (3) misguided (e.g., the substitution of cysteine residues in secretory proteins). However, when the number of desired mutants exceeds 20 or so, it becomes impractical and expensive to generate each of them individually using a separate mutagenic oligonucleotide. Methods have therefore been devised to use degenerate pools of oligonucleotides to create large populations of mutants in a single round of site-directed mutagenesis. These degenerate pools of oligonucleotides contain a mixture of normal and abnormal bases at each position in the sequence at which a mutagenic event is desired. In the remainder of this section, we present guidelines for ways in which these populations of clustered mutations can be efficiently generated using degenerate pools of mutagenic oligonucleotides.

Use of Degenerate Pools of Mutagenic Oligonucleotides

1. Pools of degenerate single-stranded oligonucleotides can be used only when the target amino acids are clustered. If all of the codons that are to be altered lie within a short stretch of contiguous nucleotides, a pool of degenerate single-stranded mutagenic oligonucleotides can be used as mismatched primers on single-stranded DNA templates to generate the corresponding set of mutants. However, the mutants cannot generally be distinguished from the original wild-type DNA by the standard method of screening by hybridization. In most cases, the mutagenic oligonucleotides are so long that there is no practical difference in stability between mismatched and perfect hybrids. Even if the oligonucleotides are sufficiently short (≤ 20 nucleotides in length), the pool usually contains many different members, each of which has different hybridization characteristics. It is therefore extremely difficult, if not impossible, to devise hybridization conditions that will distinguish all possible mutant sequences from the original wild-type sequences. This type of mutagenesis is therefore best carried out using the Kunkel system (see pages 15.74–15.79), which selects strongly against bacteriophages generated by replication of the original wild-type (+) strand of DNA. Mutants are then identified by picking individual plaques blindly and sequencing the relevant section of single-stranded bacteriophage DNA.

2. An alternative method is to generate pools of mutants by "cassette mutagenesis," a technique that involves replacing the wild-type sequence with synthetic double-stranded oligonucleotides (see, e.g., McNeil and Smith 1985; Wells et al. 1985; Derbyshire et al. 1986; Hill et al. 1986, 1987; Hutchison et al. 1986; Bedwell et al. 1989). Since cassette mutagenesis was first introduced (Matteucci and Heyneker 1983), several variations have been described, each of which has advantages under particular circumstances. However, all of these techniques suffer from a common drawback—the necessity for unique restriction sites at both ends of the cassette. Because these restriction sites are required to shuttle the synthetic double-stranded oligonucleotide into the correct location, they cannot occur anywhere else in either the plasmid vector or the segment of the wild-type gene that it carries. Furthermore, to ensure that the cassette is inserted in the correct orientation, the cassette should carry different restriction sites at each end. Because naturally occurring restriction sites hardly ever fulfill these criteria, it is usually necessary to carry out one or more rounds of site-directed mutagenesis to create suitable restriction sites at the appropriate locations in the wild-type gene. If the introduction of these sites changes the amino acid sequence encoded by the gene, it is necessary to determine whether the resulting protein displays wild-type characteristics. To eliminate the possibility that the phenotypes of any mutants obtained by cassette mutagenesis result from a combination of amino acid changes (i.e., changes caused by introduction of the restriction sites and by changes encoded within the cassette), it may be necessary to restore the original wild-type sequence at the restriction sites.

   Three different methods are currently used to generate double-stranded

oligonucleotide cassettes. In the first method (McNeil and Smith 1985) (see Figure 15.9A), two separate sets of oligonucleotides are synthesized that are complementary to the opposite strands of the target DNA. One of these sets consists of a single species of oligonucleotide that is exactly complementary to the sequence of one of the strands of the wild-type target DNA. The other set consists of a degenerate pool of oligonucleotides that is complementary to the opposite strand and that carries the desired mutations. These sets of complementary oligonucleotides are then mixed under conditions that will allow mismatched hybrids to form. If the complementary oligonucleotides have been designed to yield double-stranded cassettes that carry the appropriate protruding termini, they can be inserted directly into a recombinant plasmid in place of the homologous wild-type sequence. Alternatively, cohesive termini can be created by digesting double-stranded blunt-ended cassettes with the appropriate restriction enzymes. The mismatches in the recombinant plasmids are repaired in vivo, after the recombinant plasmids have been introduced into competent bacteria. Subsequent replication of the plasmid DNA and segregation into daughter cells allows clones to be isolated that are derived from each DNA strand of the plasmid originally used for transformation. In this method and the one that follows, the plasmids isolated from individual colonies of transformed bacteria are occasionally heterogeneous, suggesting that segregation of the plasmids is sometimes incomplete. This problem can be solved by retransforming competent bacteria with plasmid DNAs extracted from pooled primary transformants. However, in this first method of cassette mutagenesis, the frequency of mutation can never exceed 50% because only one half of the progeny plasmids are derived from the mutagenized strand.

In the second method (see Figure 15.9B), the frequency of mutation is increased by using cassettes in which the complementary strands both consist of mixed-sequence oligonucleotides (Wells et al. 1985). Because each of these strands gives rise to progeny plasmids, the mutation rate can be raised to greater than 50% (Makris et al. 1988).

In the third method (see Figure 15.9C), degenerate pools of single-stranded oligonucleotides are converted to a blunt-ended double-stranded form by mutually primed synthesis (Oliphant et al. 1986; Hill et al. 1987). Two degenerate pools of oligonucleotides are synthesized that are complementary to the same strand of the target DNA. However, the members of one pool carry sequences at their 3′ termini that are complementary to sequences at the 3′ termini of oligonucleotides in the second pool. Usually, these complementary sequences are palindromic and correspond to the restriction site that marks one end of the cassette. The oligonucleotides in the two pools are then annealed to form partial hybrids that can be converted to blunt-ended double-stranded DNA by the Klenow fragment of *E. coli* DNA polymerase I. The products of this reaction are tail-to-tail dimers. Unit-length cassettes are generated by digesting the dimers with the appropriate restriction enzymes.

The major advantage of the third method is that the unit-length cassette consists of perfect homoduplexes. Any potential bias that occurs during mismatch repair in vivo is therefore avoided, and there is no loss of

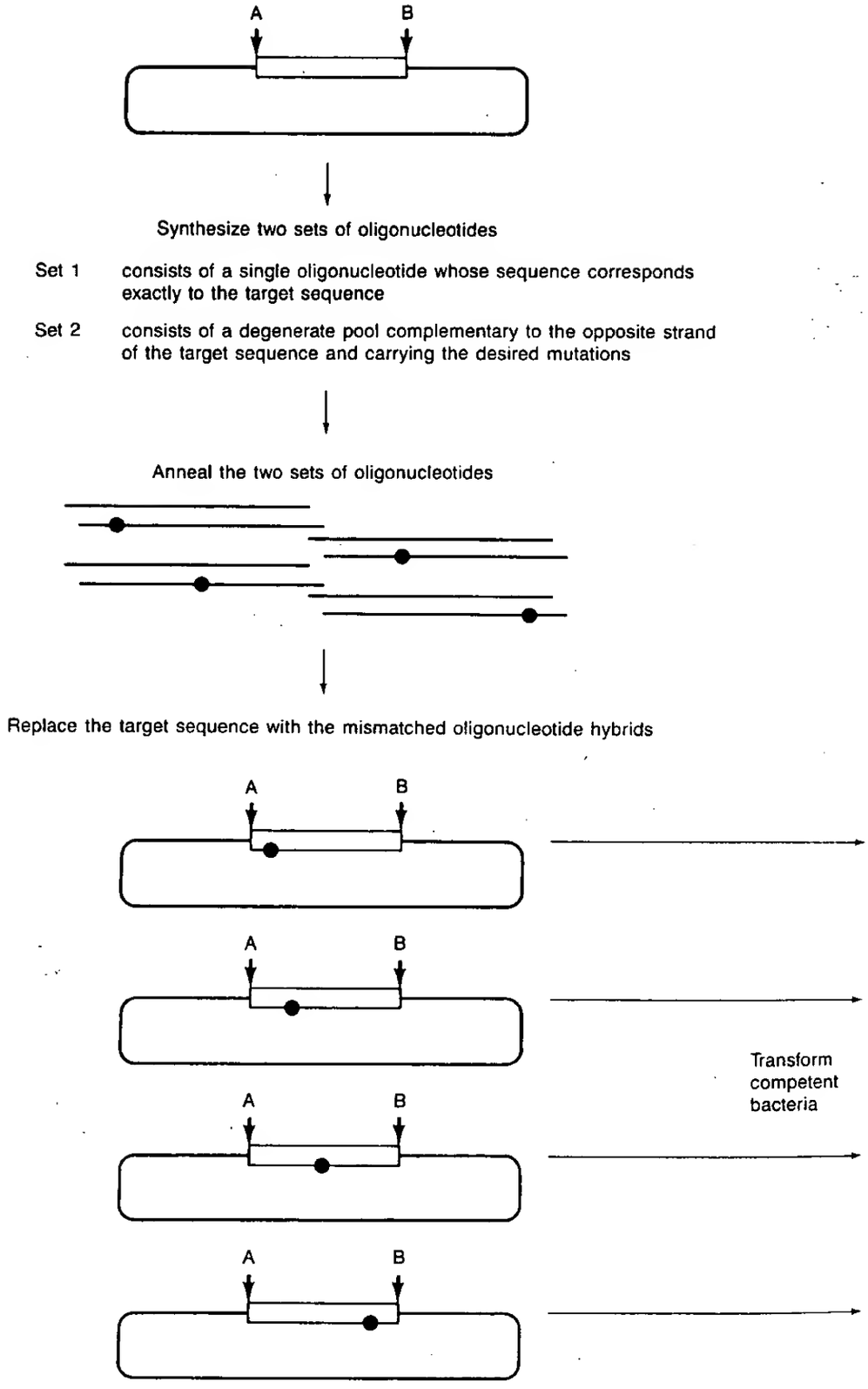
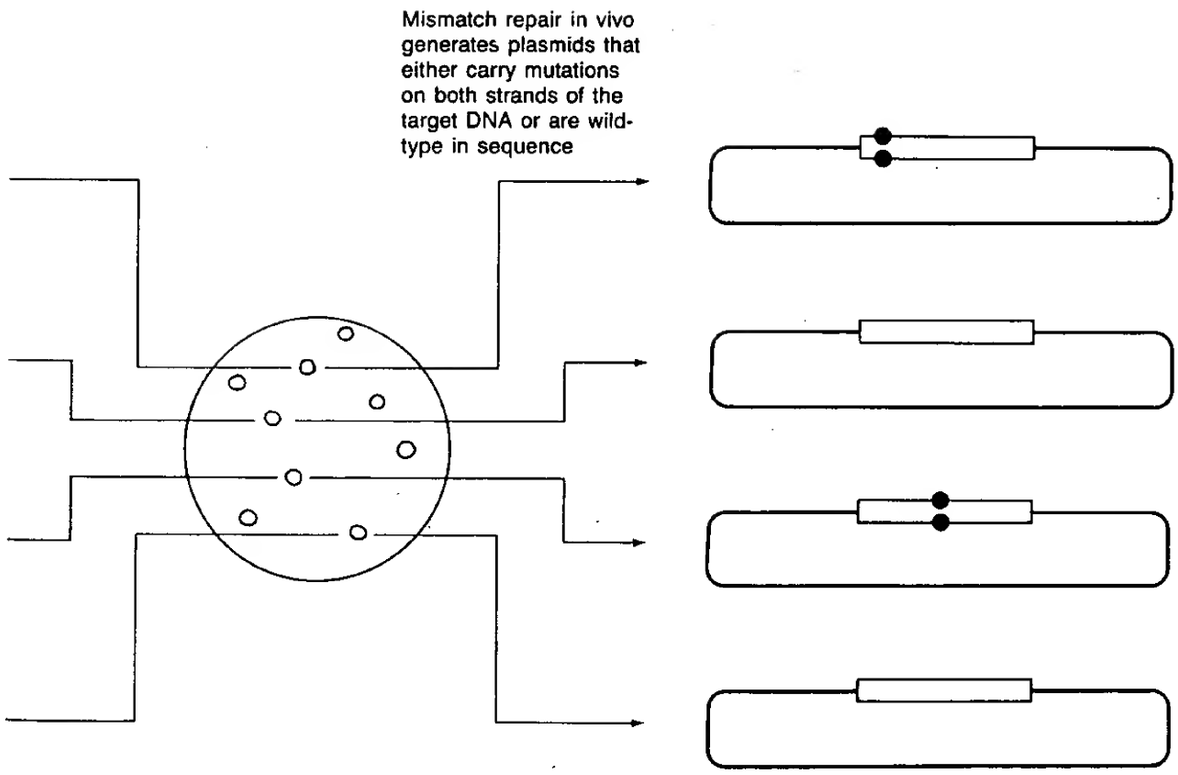


FIGURE 15.9A
Cassette mutagenesis using a single mixed-sequence oligonucleotide and repair of mismatches in vivo.


Mismatch repair in vivo generates plasmids that either carry mutations on both strands of the target DNA or are wild-type in sequence

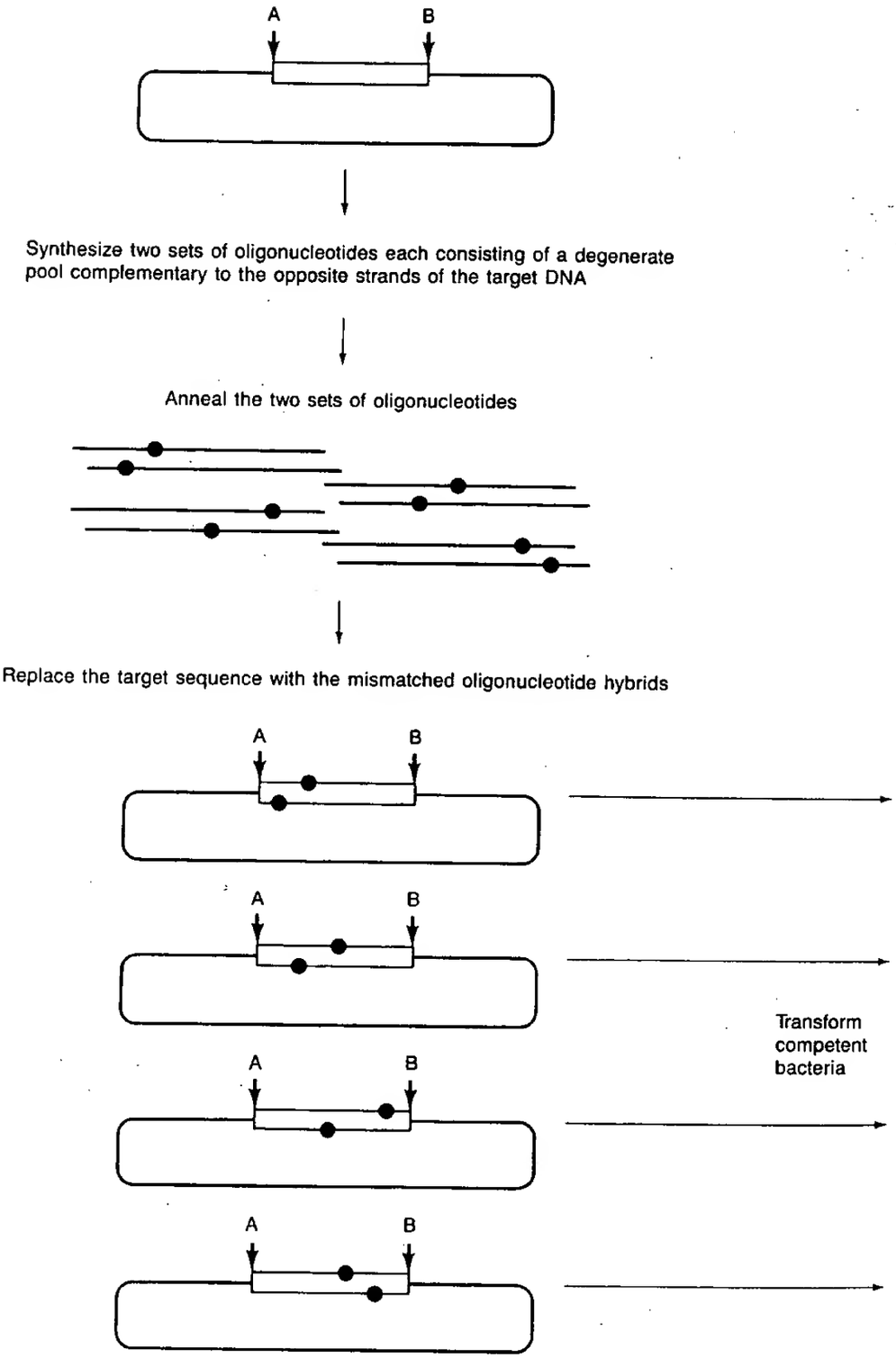
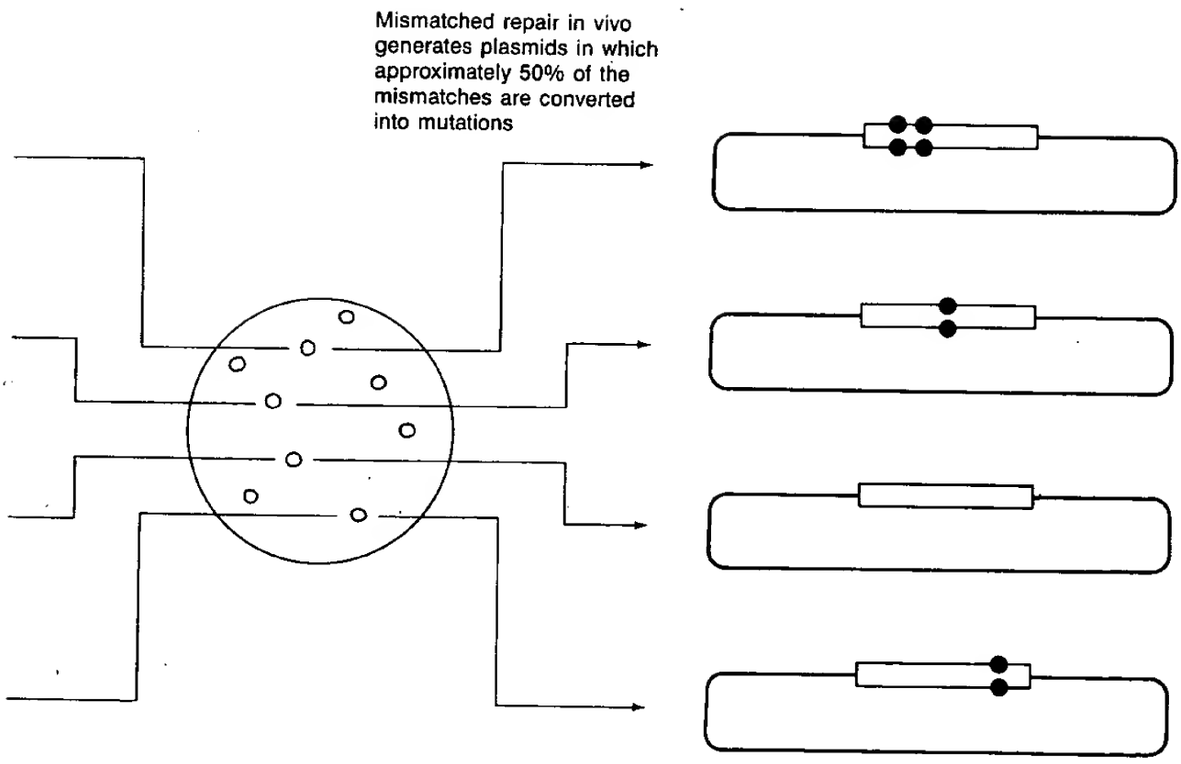


FIGURE 15.9B
Cassette mutagenesis using two complementary mixed-sequence oligonucleotides and repair of mismatches in vivo.


Mismatched repair in vivo
generates plasmids in which
approximately 50% of the
mismatches are converted
into mutations

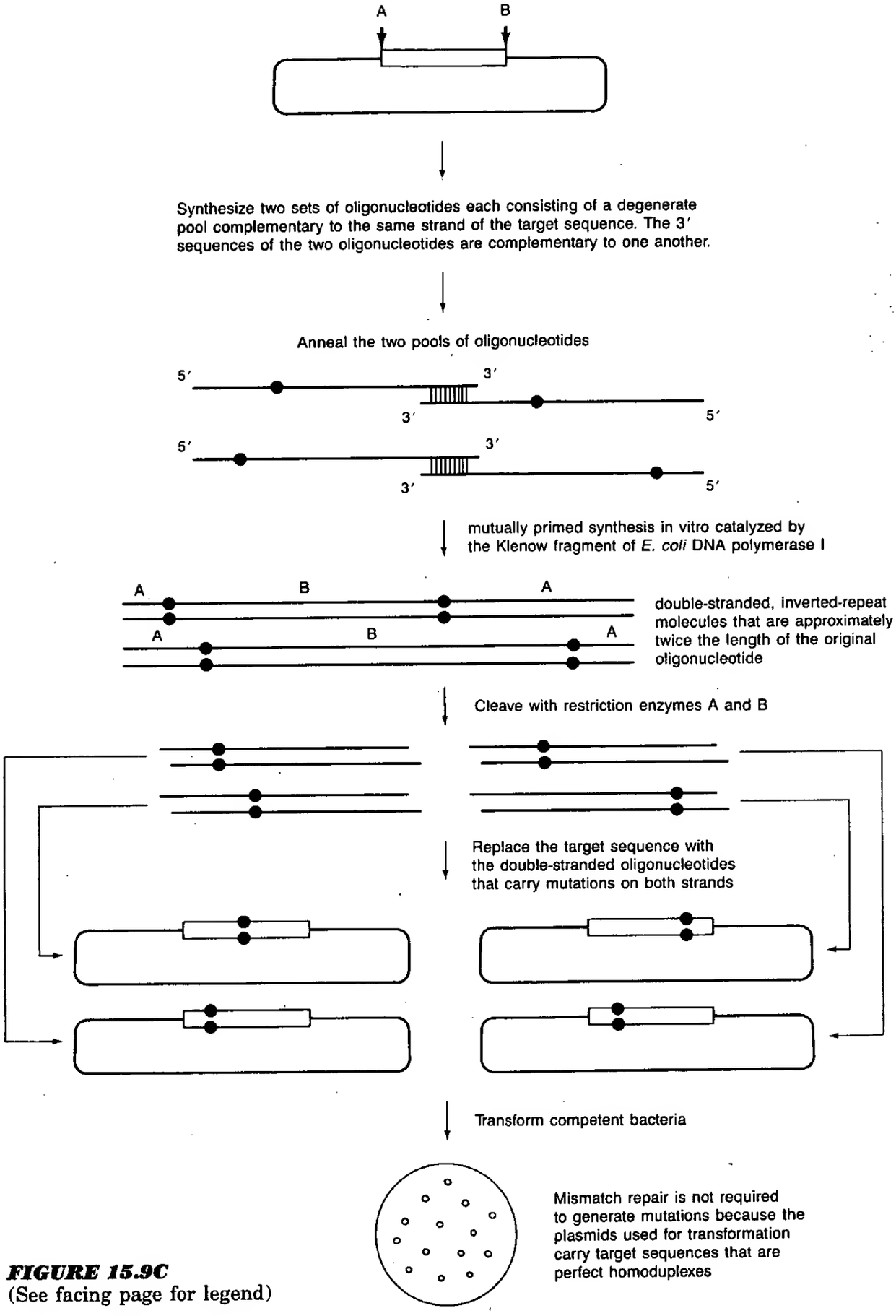


FIGURE 15.9C
(See facing page for legend)

mutants because of correction to wild-type sequences. Finally, because segregation is not required, primary bacterial transformants contain pure plasmid populations that can be analyzed directly. For these reasons, this is currently the method of choice for creating mutations at many sites within a defined region of DNA.

3. Ideally, each member of a degenerate pool of oligonucleotides should contain one nucleotide change per target sequence. In practice, however, because the oligonucleotide pools are generated in a mixed synthetic reaction, the best that can be achieved is an *average* of one altered base per target sequence. At each cycle, therefore, there is a chance that either a normal or an altered base will be incorporated into a growing oligonucleotide chain. The mutation frequency at any given site depends on the relative concentrations of the different nucleotide precursors that are provided at a particular cycle in the synthetic reaction.

   The fraction of oligonucleotides that contain nucleotide changes follows a binomial distribution that can be predicted from the following equation (McNeil and Smith 1985; Makris et al. 1988):

   $$F(P) = n!P^{n-r}(1-P)^r/(n-r)!r!$$

   where $F(P)$ is the fraction of the population whose sequence contains r random base changes over a target sequence of n consecutive bases, P is the probability of any given nucleotide being unchanged, and $(1-P)$ is the probability of any given nucleotide being changed. For example, when the length of the target sequence is 20 and the mixture of precursors supplied at every cycle contains 95% of the "normal" nucleotides and 5% of the "altered" nucleotides, the fraction of oligonucleotides that contain one altered nucleotide ($r = 1$) can be calculated as follows:

   $$F(P) = 20!(0.95)^{19}(0.05)^1/19!1!$$
   $$= 0.38$$

   Similarly, 36% of the oligonucleotides in the pool will contain no alteration in nucleotide sequence; 19% of the oligonucleotides will contain two changes, and 7% will contain more than two changes.

4. The types of mutations created by degenerate pools of oligonucleotides depend on the precursors that are supplied at each round of the synthetic cycle. In the example discussed above, not more than 5% of the precursors provided at each round of synthesis can contain "abnormal" nucleotides. Within this 5%, however, the ratio of the three abnormal bases can be altered according to the needs of the particular experiment. Because transversions are usually more useful than transitions, many workers increase the proportion of abnormal bases that will generate transversions at the expense of abnormal bases that will cause transitions.

5. The termini of the oligonucleotides should not be mutagenized because they will be needed to insert the cassette into the appropriate plasmid. If

FIGURE 15.9C
Cassette mutagenesis using two partially overlapping mixed-sequence oligonucleotides and complementary strand synthesis in vitro.

cohesive termini are to be generated by cleaving the double-stranded cassettes with restriction enzymes, three extra nucleotides should be added to each end of the mutagenic oligonucleotide. These extensions increase the efficiency of digestion with restriction enzymes.

6. The frequency with which mutants are obtained at any particular position decreases as the length of the mutagenic oligonucleotides in the degenerate pool increases. Because individual mutants are recovered by random sampling, it is improbable that all possible mutations will be isolated when the size of the potential pool is large. Under these circumstances, "missing" clones that carry particularly interesting mutations can usually be identified by hybridization to specifically designed oligonucleotide probes.

Finally, it is worth remembering that oligonucleotide-mediated mutagenesis is not the only method that can be used to saturate segments of cloned DNA with mutations. Several of the other techniques that are available are discussed below.

Treatment of Double-stranded DNA with Chemical Mutagens

The simplest method of localized random mutagenesis is to react a short fragment of double-stranded DNA with a chemical mutagen such as nitrous acid or hydroxylamine and to clone the population of mutagenized fragments into a recombinant plasmid that carries the remainder of the wild-type gene. Recombinant plasmids carrying mutations that generate a novel phenotype can be recognized by appropriate tests. For example, a temperature-sensitive mutation constructed in a gene coding for a mammalian protein might be recognized by immunofluorescent staining of mammalian cells that had been transfected with the appropriate plasmid incubated at permissive and non-permissive temperatures. Recombinant plasmids carrying a mutation that does not give rise to an easily assayed phenotype must be identified by DNA sequencing of random clones. Unfortunately, the frequency at which mutants are recovered by this method is unacceptedly low (Chu et al. 1979; Solnick 1981; Busby et al. 1982; Kadonaga and Knowles 1985). Furthermore, because chemical mutagens react with bases in double-stranded DNA in a highly specific manner, only a limited spectrum of mutations is recovered. For these reasons, this method is no longer in widespread use.

Treatment of Single-stranded DNA with Sodium Bisulfite

In the original descriptions of this protocol, circular double-stranded plasmid DNA was nicked at a random site with pancreatic DNAase I in the presence of ethidium bromide (Greenfield et al. 1975; Shortle and Botstein 1983). The nick was then converted to a gap by digestion with exonuclease III, and the resulting gapped double-stranded molecule was exposed at slightly acid pH to sodium bisulfite (1–3 M), which caused deamination of cytosine to uracil. After transformation of bacteria, replication of the mutagenized DNA led to replacement of the original C:G base pair with a T:A base pair. Recently, the efficiency of this type of mutagenesis has been improved by carrying out deamination on gapped duplexes of bacteriophage M13 recombinants in which the target DNA is exposed in a single-stranded form (Pine and Huang 1987). After mutagenesis, the DNA is transfected into an *ung*$^-$ strain of *E. coli* that is unable to remove the newly generated uracil residues. Although the procedure results in highly efficient mutagenesis of a defined segment of DNA, it generates only transition mutations in which a purine replaces a purine on one strand of DNA and a pyrimidine replaces a pyrimidine on the other. Unfortunately, mutations of this type generally result in conservative substitutions of amino acids. Thus, the range of mutants that are obtained is often too narrow to allow a comprehensive analysis of a particular segment of a protein (Shortle and Nathans 1978; DiMaio and Nathans 1980; Peden and Nathans 1982).

Treatment of Single-stranded DNA with Chemicals That Damage All Four Bases

In this method (Myers et al. 1985a), single-stranded DNA of a recombinant M13 bacteriophage is exposed under defined conditions to chemicals (nitrous acid, formic acid, and hydrazine) that modify bases in single-stranded DNA without breaking the phosphodiester backbone (see Chapter 13). After removal of the chemicals, a universal sequencing primer and avian reverse transcriptase are used to synthesize the complementary strand of DNA. When the polymerase encounters damaged bases in the template strand, it incorporates nucleotides essentially at random. Because all possible nucleotides can be incorporated at a single position, there is a 75% probability of mutation at every site of damage. Furthermore, because tranversions are generated twice as frequently as transitions, the resulting mutations generate proteins with a wide spectrum of amino acid changes. After the extension reaction is completed, the double-stranded target fragment is excised and recloned into an appropriate vector. Mutants can be identified directly by DNA sequencing of random clones.

The major problem with this method is the frequency with which useful mutations can be isolated. To prevent the formation of unacceptable numbers of multiple mutants, it is necessary to limit carefully the length of time the single-stranded DNA is exposed to damaging chemicals. However, this means that many of the template strands escape modification altogether. Therefore, the best that can be achieved by this method is a frequency of single mutations of 10–15%. This problem can sometimes be alleviated by using denaturing gradient gel electrophoresis to purify fragments of DNA that carry mutations (Myers et al. 1985a,b). However, this technique is by no means simple, and it requires the attachment of the mutagenized DNA to special vectors equipped with GC clamps (Myers et al. 1985c). Because of these problems, this method of mutagenesis has so far not found widespread acceptance.

Misincorporation of Nucleotides by DNA Polymerases

Point mutations can be introduced into double-stranded DNA by incorporating base analogs with various types of DNA polymerases. For example, Shortle and his coworkers (Shortle et al. 1982; Shortle and Lin 1985) incubated gapped DNA in the presence of *E. coli* DNA polymerase I and only one of the four α-thiophosphate dNTPs. Thiophosphate dNTPs are efficiently incorporated by the polymerase but are not effectively removed by its $3' \rightarrow 5'$ editing function. The incorrect base is thus incorporated at a high frequency, and the remainder of the gap is then filled in a second polymerization reaction carried out in the presence of all four of the normal dNTPs. All types of base substitutions have been obtained with this method using each of the four α-thiophosphate dNTPs in separate repair reactions.

Another misincorporation method uses AMV reverse transcriptase, which is deficient in a $3' \rightarrow 5'$ exonuclease activity (Zakour and Loeb 1982). In this case, conventional dNTPs are used to synthesize DNA from an upstream primer. Base analogs are then incorporated in the region of interest.

The major problem of these and other misincorporation methods is the difficulty in creating populations of template molecules in which the 3′ hydroxyl terminus of the growing strand is located at random positions throughout the region of interest. Although this can in theory be achieved by a number of different methods (e.g., controlled nick translation with *E. coli* DNA polymerase I or controlled digestion of double-stranded DNA with exonuclease III), the routine generation of large numbers of mutations at random sites has proved to be difficult in practice. Success requires careful characterization of the reagents involved, meticulous establishment of optimal reaction conditions, and many trial experiments.

In summary, whereas methods (discussed earlier in this chapter) to introduce single mutations in cloned DNA are now well-established, techniques to saturate defined regions with mutations are less satisfactory. Using chemical mutagenesis, the rate of production of single mutations is low and/or the mutations themselves are of limited interest. Using misincorporation of base analogs, it is difficult to direct the mutations to the region of interest. However, it seems likely that at least some of these problems will be solved during the next few years, for example, by incorporating base changes into DNA synthesized in polymerase chain reactions or by advances in DNA chemistry that will facilitate the synthesis of mutagenized DNAs of extended length. Until then, we recommend using degenerate pools of synthetic oligonucleotides. In contrast to the other methods, the mutations can be precisely designed by the experimenter and can be focused in a defined region of DNA. Even with these limitations, the amount of work involved in isolating and characterizing a comprehensive set of mutants remains very large. In this branch of molecular cloning, therefore, it is especially important to weigh the potential scientific rewards against the commitment of time and personnel that the project will certainly consume.